information and human values kenneth r fleischmann

#information ethics #human values #kenneth r fleischmann #digital ethics #information science

Explore the intersection of information and human values as examined by Kenneth R. Fleischmann. This analysis delves into the ethical considerations surrounding information access, use, and dissemination in the digital age, considering the impact on individual autonomy, social justice, and the preservation of fundamental human values.

Information and Human Values

This book seeks to advance our understanding of the relationship between information and human values by synthesizing the complementary but typically disconnected threads in the literature, reflecting on my 15 years of research on the relationship between information and human values, advancing our intellectual understanding of the key facets of this topic, and encouraging further research to continue exploring this important and timely research topic. The book begins with an explanation of what human values are and why they are important. Next, three distinct literatures on values, information, and technology are analyzed and synthesized, including the social psychology literature on human values, the information studies literature on the core values of librarianship, and the human-computer interaction literature on value-sensitive design. After that, three detailed case studies are presented based on reflections on a wide range of research studies. The first case study focuses on the role of human values in the design and use of educational simulations. The second case study focuses on the role of human values in the design and use of computational models. The final case study explores human values in communication via, about, or using information technology. The book concludes by laying out a values and design cycle for studying values in information and presenting an agenda for further research.

Information and Human Values

This book seeks to advance our understanding of the relationship between information and human values by synthesizing the complementary but typically disconnected threads in the literature, reflecting on my 15 years of research on the relationship between information and human values, advancing our intellectual understanding of the key facets of this topic, and encouraging further research to continue exploring this important and timely research topic. The book begins with an explanation of what human values are and why they are important. Next, three distinct literatures on values, information, and technology are analyzed and synthesized, including the social psychology literature on human

values, the information studies literature on the core values of librarianship, and the human-computer interaction literature on value-sensitive design. After that, three detailed case studies are presented based on reflections on a wide range of research studies. The first case study focuses on the role of human values in the design and use of educational simulations. The second case study focuses on the role of human values in the design and use of computational models. The final case study explores human values in communication via, about, or using information technology. The book concludes by laying out a values and design cycle for studying values in information and presenting an agenda for further research.

Social Informatics

Social Informatics: Past, Present and Future is a collection of twelve papers that provides a state-of-the-art review of 21st century social informatics. Two papers review the history of social informatics, and show that its intellectual roots can be found in the late 1970s and early ’80s and that it emerged in several different locations around the world before it coalesced in the US in the mid-1990s. The evolution of social informatics is described under four periods: foundational work, development and expansion, a robust period of coherence, and a period of diversification that continues today. Five papers provide a view of the breadth and depth of contemporary social informatics, demonstrating the diversity of theoretical and methodological approaches that can be used. A further five papers explore the future of social informatics and offer provocative and disparate visions of its trajectory, ranging from arguments for a new philosophical grounding for social informatics, to calls for a social informatics based on practice thinking and materiality. This book presents a view of SI that emphasizes the core relationship among people, ICT and organizational and social life from a perspective that integrates aspects of social theory and demonstrates clearly that social informatics has never been a more necessary research endeavor than it is now.

Information in Contemporary Society

This book constitutes the proceedings of the 14th International Conference on Information in Contemporary Society, iConference 2019, held in Washington, DC, USA, in March/April 2019. The 44 full papers and 33 short papers presented in this volume were carefully reviewed and selected from 133 submitted full papers and 88 submitted short papers. The papers are organized in the following topical sections: Scientific work and data practices; methodological concerns in (big) data research; concerns about “smart” interactions and privacy; identity questions in online communities; measuring and tracking scientific literature; limits and affordances of automation; collecting data about vulnerable populations; supporting communities through public libraries and infrastructure; information behaviors in academic environments; data-driven storytelling and modeling; online activism; digital libraries, curation and preservation; social-media text mining and sentiment analysis; data and information in the public sphere; engaging with multi-media content; understanding online behaviors and experiences; algorithms at work; innovation and professionalization in technology communities; information behaviors on Twitter; data mining and NLP; informing technology design through offline experiences; digital tools for health management; environmental and visual literacy; and addressing social problems in iSchool research.

Information Architecture

Information Architecture is about organizing and simplifying information, designing and integrating information spaces/systems, and creating ways for people to find and interact with information content. Its goal is to help people understand and manage information and make the right decisions accordingly. This updated and revised edition of the book looks at integrated information spaces in the web context and beyond, with a focus on putting theories and principles into practice. In the ever-changing social, organizational, and technological contexts, information architects not only design individual information spaces (e.g., websites, software applications, and mobile devices), but also tackle strategic aggregation and integration of multiple information spaces across websites, channels, modalities, and platforms. Not only do they create predetermined navigation pathways, but they also provide tools and rules for people to organize information on their own and get connected with others. Information architects work with multi-disciplinary teams to determine the user experience strategy based on user needs and business goals, and make sure the strategy gets carried out by following the user-centered design (UCD) process via close collaboration with others. Drawing on the authors’ extensive experience as HCI researchers, User Experience Design practitioners, and Information Architecture instructors, this book

provides a balanced view of the IA discipline by applying theories, design principles, and guidelines to IA and UX practices. It also covers advanced topics such as iterative design, UX decision support, and global and mobile IA considerations. Major revisions include moving away from a web-centric view toward multi-channel, multi-device experiences. Concepts such as responsive design, emerging design principles, and user-centered methods such as Agile, Lean UX, and Design Thinking are discussed and related to IA processes and practices.

Predicting Information Retrieval Performance

Information Retrieval performance measures are usually retrospective in nature, representing the effectiveness of an experimental process. However, in the sciences, phenomena may be predicted, given parameter values of the system. After developing a measure that can be applied retrospectively or can be predicted, performance of a system using a single term can be predicted given several different types of probabilistic distributions. Information Retrieval performance can be predicted with multiple terms, where statistical dependence between terms exists and is understood. These predictive models may be applied to realistic problems, and then the results may be used to validate the accuracy of the methods used. The application of metadata or index labels can be used to determine whether or not these features should be used in particular cases. Linguistic information, such as part-of-speech tag information, can increase the discrimination value of existing terminology and can be studied predictively. This work provides methods for measuring performance that may be used predictively. Means of predicting these performance measures are provided, both for the simple case of a single term in the query and for multiple terms. Methods of applying these formulae are also suggested.

Information Communication

This book introduces fundamentals of information communication. At first, concepts and characteristics of information and information communication are summarized. And then five classic models of information communication are introduced. The mechanisms and fundamental laws of the information transmission process are also discussed. In order to realize information communication, impediments in information communication process are identified and analyzed. For the purpose of investigating implications of Internet information communication, patterns and characteristics of information communication in the Internet and Web 2.0 environment are also analyzed. In the end, case studies are provided for readers to understand the theory.

Fuzzy Information Retrieval

Information retrieval used to mean looking through thousands of strings of texts to find words or symbols that matched a user's query. Today, there are many models that help index and search more effectively so retrieval takes a lot less time. Information retrieval (IR) is often seen as a subfield of computer science and shares some modeling, applications, storage applications and techniques, as do other disciplines like artificial intelligence, database management, and parallel computing. This book introduces the topic of IR and how it differs from other computer science disciplines. A discussion of the history of modern IR is briefly presented, and the notation of IR as used in this book is defined. The complex notation of relevance is discussed. Some applications of IR is noted as well since IR has many practical uses today. Using information retrieval with fuzzy logic to search for software terms can help find software components and ultimately help increase the reuse of software. This is just one practical application of IR that is covered in this book. Some of the classical models of IR is presented as a contrast to extending the Boolean model. This includes a brief mention of the source of weights for the various models. In a typical retrieval environment, answers are either yes or no, i.e., on or off. On the other hand, fuzzy logic can bring in a "degree of" match, vs. a crisp, i.e., strict match. This, too, is looked at and explored in much detail, showing how it can be applied to information retrieval. Fuzzy logic is often times considered a soft computing application and this book explores how IR with fuzzy logic and its membership functions as weights can help indexing, querying, and matching. Since fuzzy set theory and logic is explored in IR systems, the explanation of where the fuzz is ensues. The concept of relevance feedback, including pseudorelevance feedback is explored for the various models of IR. For the extended Boolean model, the use of genetic algorithms for relevance feedback is delved into. The concept of query expansion is explored using rough set theory. Various term relationships is modeled and presented, and the model extended for fuzzy retrieval. An example using the UMLS terms is also presented. The model is also extended for term relationships beyond synonyms. Finally, this book looks

at clustering, both crisp and fuzzy, to see how that can improve retrieval performance. An example is presented to illustrate the concepts.

Dynamic Information Retrieval Modeling

Big data and human-computer information retrieval (HCIR) are changing IR. They capture the dynamic changes in the data and dynamic interactions of users with IR systems. A dynamic system is one which changes or adapts over time or a sequence of events. Many modern IR systems and data exhibit these characteristics which are largely ignored by conventional techniques. What is missing is an ability for the model to change over time and be responsive to stimulus. Documents, relevance, users and tasks all exhibit dynamic behavior that is captured in data sets typically collected over long time spans and models need to respond to these changes. Additionally, the size of modern datasets enforces limits on the amount of learning a system can achieve. Further to this, advances in IR interface, personalization and ad display demand models that can react to users in real time and in an intelligent, contextual way. In this book we provide a comprehensive and up-to-date introduction to Dynamic Information Retrieval Modeling, the statistical modeling of IR systems that can adapt to change. We define dynamics, what it means within the context of IR and highlight examples of problems where dynamics play an important role. We cover techniques ranging from classic relevance feedback to the latest applications of partially observable Markov decision processes (POMDPs) and a handful of useful algorithms and tools for solving IR problems incorporating dynamics. The theoretical component is based around the Markov Decision Process (MDP), a mathematical framework taken from the field of Artificial Intelligence (AI) that enables us to construct models that change according to sequential inputs. We define the framework and the algorithms commonly used to optimize over it and generalize it to the case where the inputs aren't reliable. We explore the topic of reinforcement learning more broadly and introduce another tool known as a Multi-Armed Bandit which is useful for cases where exploring model parameters is beneficial. Following this we introduce theories and algorithms which can be used to incorporate dynamics into an IR model before presenting an array of state-of-the-art research that already does, such as in the areas of session search and online advertising. Change is at the heart of modern Information Retrieval systems and this book will help equip the reader with the tools and knowledge needed to understand Dynamic Information Retrieval Modeling.

Building a Better World with Our Information

Part 1 in "The Future of" series covers the fundamentals of personal information management (PIM) and then explores the seismic shift, already well underway, toward a world where our information is always at hand (thanks to our devices) and "forever" on the web. Part 2, "Transforming Technologies to Manage Our Information," provides a more focused look at technologies for managing information. The opening chapter discusses "natural interface" technologies of input/output to free us from keyboard, screen, and mouse. Successive chapters then explore technologies to save, search, and structure our information. A concluding chapter introduces the possibility that we may see dramatic reductions in the "clerical tax" we pay as we work with our information. Focus in this concluding Part 3 to the series shifts to the practical and to the near future. What can we do, now or soon, to manage our information better? And, as we do so, how might we build a better world? Part 3 is in three chapters: Chapter 10. Group Information Management and the Social Fabric in PIM. How do we preserve and promote our PIM practices as we interact with others at home, at school, at work, at play and in wider, even global, communities? Chapter 11. PIM by Design. What principles guide us? How can developers build better tools for PIM? How can the rest of us make better use of the tools we already have? Chapter 12. To Each of Us, Our Own concludes with an exploration of the ways each of us, individually, can develop better practices for the management of our information in service of the lives we wish to live and toward a better world we all must share.

Simulating Information Retrieval Test Collections

Simulated test collections may find application in situations where real datasets cannot easily be accessed due to confidentiality concerns or practical inconvenience. They can potentially support Information Retrieval (IR) experimentation, tuning, validation, performance prediction, and hardware sizing. Naturally, the accuracy and usefulness of results obtained from a simulation depend upon the fidelity and generality of the models which underpin it. The fidelity of emulation of a real corpus is likely to be limited by the requirement that confidential information in the real corpus should not be able to be extracted from the emulated version. We present a range of methods exploring trade-offs between

emulation fidelity and degree of preservation of privacy. We present three different simple types of text generator which work at a micro level: Markov models, neural net models, and substitution ciphers. We also describe macro level methods where we can engineer macro properties of a corpus, giving a range of models for each of the salient properties: document length distribution, word frequency distribution (for independent and non-independent cases), word length and textual representation, and corpus growth. We present results of emulating existing corpora and for scaling up corpora by two orders of magnitude. We show that simulated collections generated with relatively simple methods are suitable for some purposes and can be generated very quickly. Indeed it may sometimes be feasible to embed a simple lightweight corpus generator into an indexer for the purpose of efficiency studies. Naturally, a corpus of artificial text cannot support IR experimentation in the absence of a set of compatible queries. We discuss and experiment with published methods for query generation and query log emulation. We present a proof-of-the-pudding study in which we observe the predictive accuracy of efficiency and effectiveness results obtained on emulated versions of TREC corpora. The study includes three open-source retrieval systems and several TREC datasets. There is a trade-off between confidentiality and prediction accuracy and there are interesting interactions between retrieval systems and datasets. Our tentative conclusion is that there are emulation methods which achieve useful prediction accuracy while providing a level of confidentiality adequate for many applications. Many of the methods described here have been implemented in the open source project SynthaCorpus, accessible at: https://bitbucket.org/davidhawking/synthacorpus/

Researching Serendipity in Digital Information Environments

Chance, luck, and good fortune are the usual go-to descriptors of serendipity, a phenomenon aptly often coupled with famous anecdotes of accidental discoveries in engineering and science in modern history such as penicillin, Teflon, and Post-it notes. Serendipity, however, is evident in many fields of research, in organizations, in everyday life—and there is more to it than luck implies. While the phenomenon is strongly associated with in person interactions with people, places, and things, most attention of late has focused on its preservation and facilitation within digital information environments. Serendipity's association with unexpected, positive user experiences and outcomes has spurred an interest in understanding both how current digital information environments support serendipity and how novel approaches may be developed to facilitate it. Research has sought to understand serendipity, how it is manifested in people's personality traits and behaviors, how it may be facilitated in digital information environments such as mobile applications, and its impacts on an individual, an organizational, and a wider level. Because serendipity is expressed and understood in different ways in different contexts, multiple methods have been used to study the phenomenon and evaluate digital information environments that may support it. This volume brings together different disciplinary perspectives and examines the motivations for studying serendipity, the various ways in which serendipity has been approached in the research, methodological approaches to build theory, and how it may be facilitated. Finally, a roadmap for serendipity research is drawn by integrating key points from this volume to produce a framework for the examination of serendipity in digital information environments.

Video Structure Meaning

For over a century, motion pictures have entertained us, occasionally educated us, and even served a few specialized fields of study. Now, however, with the precipitous drop in prices and increase in image quality, motion pictures are as widespread as paperback books and postcards once were. Yet, theories and practices of analysis for particular genres and analytical stances, definitions, concepts, and tools that span platforms have been wanting. Therefore, we developed a suite of tools to enable close structural analysis of the time-varying signal set of a movie. We take an information-theoretic approach (message is a signal set) generated (coded) under various antecedents (sent over some channel) decoded under some other set of antecedents. Cultural, technical, and personal antecedents might favor certain message-making systems over others. The same holds true at the recipient end-yet, the signal set remains the signal set. In order to discover how movies work-their structure and meaning-we honed ways to provide pixel level analysis, forms of clustering, and precise descriptions of what parts of a signal influence viewer behavior. We assert that analysis of the signal set across the evolution of film—from Edison to Hollywood to Brakhage to cats on social media—yields a common ontology with instantiations (responses to changes in coding and decoding antecedents).

Third Space, Information Sharing, and Participatory Design

Society faces many challenges in workplaces, everyday life situations, and education contexts. Within information behavior research, there are often calls to bridge inclusiveness and for greater collaboration, with user-centered design approaches and, more specifically, participatory design practices. Collaboration and participation are essential in addressing contemporary societal challenges, designing creative information objects and processes, as well as developing spaces for learning, and information and research interventions. The intention is to improve access to information and the benefits to be gained from that. This also applies to bridging the digital divide and for embracing artificial intelligence. With regard to research and practices within information behavior, it is crucial to consider that all users should be involved. Many information activities (i.e., activities falling under the umbrella terms of information behavior and information practices) manifest through participation, and thus, methods such as participatory design may help unfold both information behavior and practices as well as the creation of information objects, new models, and theories. Information sharing is one of its core activities. For participatory design with its value set of democratic, inclusive, and open participation towards innovative practices in a diversity of contexts, it is essential to understand how information activities such as sharing manifest itself. For information behavior studies it is essential to deepen understanding of how information sharing manifests in order to improve access to information and the use of information. Third Space is a physical, virtual, cognitive, and conceptual space where participants may negotiate, reflect, and form new knowledge and worldviews working toward creative, practical and applicable solutions, finding innovative, appropriate research methods, interpreting findings, proposing new theories, recommending next steps, and even designing solutions such as new information objects or services. Information sharing in participatory design manifests in tandem with many other information interaction activities and especially information and cognitive processing. Although there are practices of individual information sharing and information encountering, information sharing mostly relates to collaborative information behavior practices, creativity, and collective decision-making. Our purpose with this book is to enable students, researchers, and practitioners within a multi-disciplinary research field, including information studies and Human–Computer Interaction approaches, to gain a deeper understanding of how the core activity of information sharing in participatory design, in which Third Space may be a platform for information interaction, is taking place when using methods utilized in participatory design to address contemporary societal challenges. This could also apply for information behavior studies using participatory design as methodology. We elaborate interpretations of core concepts such as participatory design, Third Space, information sharing, and collaborative information behavior, before discussing participatory design methods and processes in more depth. We also touch on information behavior, information practice, and other important concepts. Third Space, information sharing, and information interaction are discussed in some detail. A framework, with Third Space as a core intersecting zone, platform, and adaptive and creative space to study information sharing and other information behavior and interactions are suggested. As a tool to envision information behavior and suggest future practices, participatory design serves as a set of methods and tools in which new interpretations of the design of information behavior studies and eventually new information objects are being initiated involving multiple stakeholders in future information landscapes. For this purpose, we argue that Third Space can be used as an intersection zone to study information sharing and other information activities, but more importantly it can serve as a Third Space Information Behavior (TSIB) study framework where participatory design methodology and processes are applied to information behavior research studies and applications such as information objects, systems, and services with recognition of the importance of situated awareness.

Exploring Context in Information Behavior

The field of human information behavior runs the gamut of processes from the realization of a need or gap in understanding, to the search for information from one or more sources to fill that gap, to the use of that information to complete a task at hand or to satisfy a curiosity, as well as other behaviors such as avoiding information or finding information serendipitously. Designers of mechanisms, tools, and computer-based systems to facilitate this seeking and search process often lack a full knowledge of the context surrounding the search. This context may vary depending on the job or role of the person; individual characteristics such as personality, domain knowledge, age, gender, perception of self, etc.; the task at hand; the source and the channel and their degree of accessibility and usability; and the relationship that the seeker shares with the source. Yet researchers have yet to agree on what context really means. While there have been various research studies incorporating context, and biennial conferences on context in information behavior, there lacks a clear definition of what context is, what its boundaries are, and what elements and variables comprise context. In this book, we look

at the many definitions of and the theoretical and empirical studies on context, and I attempt to map the conceptual space of context in information behavior. I propose theoretical frameworks to map the boundaries, elements, and variables of context. I then discuss how to incorporate these frameworks and variables in the design of research studies on context. We then arrive at a unified definition of context. This book should provide designers of search systems a better understanding of context as they seek to meet the needs and demands of information seekers. It will be an important resource for researchers in Library and Information Science, especially doctoral students looking for one resource that covers an exhaustive range of the most current literature related to context, the best selection of classics, and a synthesis of these into theoretical frameworks and a unified definition. The book should help to move forward research in the field by clarifying the elements, variables, and views that are pertinent. In particular, the list of elements to be considered, and the variables associated with each element will be extremely useful to researchers wanting to include the influences of context in their studies.

The Notion of Relevance in Information Science

Everybody knows what relevance is. It is a "ya'know" notion, concept, idea–no need to explain whatsoever. Searching for relevant information using information technology (IT) became a ubiquitous activity in contemporary information society. Relevant information means information that pertains to the matter or problem at hand—it is directly connected with effective communication. The purpose of this book is to trace the evolution and with it the history of thinking and research on relevance in information science and related fields from the human point of view. The objective is to synthesize what we have learned about relevance in several decades of investigation about the notion in information science. This book deals with how people deal with relevance—it does not cover how systems deal with relevance; it does not deal with algorithms. Spurred by advances in information retrieval (IR) and information systems of various kinds in handling of relevance, a number of basic questions are raised: But what is relevance to start with? What are some of its properties and manifestations? How do people treat relevance? What affects relevance assessments? What are the effects of inconsistent human relevance judgments on tests of relative performance of different IR algorithms or approaches? These general questions are discussed in detail.

Children’s Internet Search

Searching the Internet and the ability to competently use search engines are increasingly becoming an important part of children’s daily lives. Whether mobile or at home, children use search interfaces to explore personal interests, complete academic assignments, and have social interaction. However, engaging with search also means engaging with an ever-changing and evolving search landscape. There are continual software updates, multiple devices used to search (e.g., phones, tablets), an increasing use of social media, and constantly updated Internet content. For young searchers, this can require infinite adaptability or mean being hopelessly confused. This book offers a perspective centered on children’s search experiences as a whole instead of thinking of search as a process with separate and potentially problematic steps. Reading the prior literature with a child-centered view of search reveals that children have been remarkably consistent over time as searchers, displaying the same search strategies regardless of the landscape of search. However, no research has synthesized these consistent patterns in children’s search across the literature, and only recently have these patterns been uncovered as distinct search roles, or searcher types. Based on a four-year longitudinal study on children’s search experiences, this book weaves together the disparate evidence in the literature through the use of 9 search roles for children ages 7-15. The search role framework has a distinct advantage because it encourages adult stakeholders to design children’s search tools to support and educate children at their existing levels of search strength and deficit, rather than expecting children to adapt to a transient search landscape.

The Taxobook

This is the first volume in a series about creating and maintaining taxonomies and their practical applications, especially in search functions. In Book 1 (The Taxobook: History, Theories, and Concepts of Knowledge Organization), the author introduces the very foundations of classification, starting with the ancient Greek philosophers Plato and Aristotle, as well as Theophrastus and the Roman Pliny the Elder. They were first in a line of distinguished thinkers and philosophers to ponder the organization of the world around them and attempt to apply a structure or framework to that world. The author continues by

discussing the works and theories of several other philosophers from Medieval and Renaissance times, including Saints Aquinas and Augustine, William of Occam, Andrea Cesalpino, Carl Linnaeus, and René Descartes. In the 17th, 18th, and 19th centuries, John Locke, Immanuel Kant, James Frederick Ferrier, Charles Ammi Cutter, and Melvil Dewey contributed greatly to the theories of classification systems and knowledge organization. Cutter and Dewey, especially, created systems that are still in use today. Chapter 8 covers the contributions of Shiyali Ramamrita Ranganathan, who is considered by many to be the “father of modern library science.” He created the concept of faceted vocabularies, which are widely used—even if they are not well understood—on many e-commerce websites. Following the discussions and historical review, the author has included a glossary that covers all three books of this series so that it can be referenced as you work your way through the second and third volumes. The author believes that it is important to understand the history of knowledge organization and the differing viewpoints of various philosophers—even if that understanding is only that the differing viewpoints simply exist. Knowing the differing viewpoints will help answer the fundamental questions: Why do we want to build taxonomies? How do we build them to serve multiple points of view? Table of Contents: List of Figures / Preface / Acknowledgments / Origins of Knowledge Organization Theory: Early Philosophy of Knowledge / Saints and Traits: Realism and Nominalism / Arranging the glowers... and the Birds, and the Insects, and Everything Else: Early Naturalists and Taxonomies / The Age of Enlightenment Impacts Knowledge Theory / 18th-Century Developments: Knowledge Theory Coming to the Foreground / High Resolution: Classification Sharpens in the 19th and 20th Centuries / Outlining the World and Its Parts / Facets: An Indian Mathematician and Children’s Toys at Selfridge’s / Points of Knowledge / Glossary / End Notes / Author Biography

Social Informatics Evolving

The study of people, information, and communication technologies and the contexts in which these technologies are designed, implemented, and used has long interested scholars in a wide range of disciplines, including the social study of computing, science and technology studies, the sociology of technology, and management information systems. As ICT use has spread from organizations into the larger world, these devices have become routine information appliances in our social lives, researchers have begun to ask deeper and more profound questions about how our lives have become bound up with technologies. A common theme running through this research is that the relationships among people, technology, and context are dynamic, complex, and critically important to understand. This book explores social informatics (SI), one important and dynamic approach that researchers have used to study these complex relationships. SI is "the interdisciplinary study of the design, uses and consequences of information technology that takes into account their interaction with institutional and cultural contexts" (Kling 1998, p. 52; 1999). SI provides flexible frameworks to explore complex and dynamic socio-technical interactions. As a domain of study related largely by common vocabulary and conclusions, SI critically examines common conceptions of and expectations for technology, by providing contextual evidence. This book describes the evolution of SI research and identifies challenges and opportunities for future research. In what might be seen as an example of socio-technical "natural selection," SI emerged in six different locations during the 1980s and 1990s: Norway, Slovenia, Japan, the former Soviet Union, the UK and, last, the U.S. As SI evolved, the version popularized in the US became globally dominant. The evolution of SI is presented in five stages: emergence, foundational, expansion, coherence, and transformation. Thus, we divide SI research into five major periods: an emergence stage, when various forms of SI emerged around the globe, an early period of foundational work which grounds SI (Pre-1990s), a period of expansion (1990s), a robust period of coherence and influence by Rob Kling (2000–2005), and a period of transformation (2006–present). Following the description of the five periods we discuss the evolution throughout the periods under five sections: principles, concepts, approaches, topics, and findings. Principles refer to the overarching motivations and labels employed to describe scholarly work. Approaches describe the theories, frameworks, and models employed in analysis, emphasizing the multi-disciplinary and interdisciplinary nature of SI. Concepts include specific processes, entities, themes, and elements of discourse within a given context, revealing a shared SI language surrounding change, complexity, consequences, and social elements of technology. Topics label the issues and general domains studied within social informatics, ranging from scholarly communication to online communities to information systems. Findings from seminal SI works illustrate growing insights over time and demonstrate how repeatable explanations unify SI. In the concluding remarks, we raise questions about the possible futures of SI research.

Trustworthy Policies for Distributed Repositories

A trustworthy repository provides assurance in the form of management documents, event logs, and audit trails that digital objects are being managed correctly. The assurance includes plans for the sustainability of the repository, the accession of digital records, the management of technology evolution, and the mitigation of the risk of data loss. A detailed assessment is provided by the ISO-16363:2012 standard, "Space data and information transfer systems—Audit and certification of trustworthy digital repositories." This book examines whether the ISO specification for trustworthiness can be enforced by computer actionable policies. An implementation of the policies is provided and the policies are sorted into categories for procedures to manage externally generated documents, specify repository parameters, specify preservation metadata attributes, specify audit mechanisms for all preservation actions, specify control of preservation operations, and control preservation properties as technology evolves. An application of the resulting procedures is made to enforce trustworthiness within National Science Foundation data management plans.

Scholarly Collaboration on the Academic Social Web

Collaboration among scholars has always been recognized as a fundamental feature of scientific discovery. The ever-increasing diversity among disciplines and complexity of research problems makes it even more compelling to collaborate in order to keep up with the fast pace of innovation and advance knowledge. Along with the rapidly developing Internet communication technologies and the increasing popularity of the social web, we have observed many important developments of scholarly collaboration on the academic social web. In this book, we review the rapid transformation of scholarly collaboration on various academic social web platforms and examine how these platforms have facilitated academics throughout their research lifecycle—from forming ideas, collecting data, and authoring articles to disseminating findings. We refer to the term "academic social web platforms" in this book as a category of Web 2.0 tools or online platforms (such as CiteULike, Mendeley, Academia.edu, and ResearchGate) that enable and facilitate scholarly information exchange and participation. We will also examine scholarly collaboration behaviors including sharing academic resources, exchanging opinions, following each other's research, keeping up with current research trends, and, most importantly, building up their professional networks. Inspired by the model developed Olson et al. [2000] on factors for successful scientific collaboration, our examination of the status of scholarly collaboration on the academic social web has four emphases: technology readiness, coupling work, building common ground, and collaboration readiness. Finally, we talk about the insights and challenges of all these online scholarly collaboration activities imposed on the research communities who are engaging in supporting online scholarly collaboration. This book aims to help researchers and practitioners understand the development of scholarly collaboration on the academic social web, and to build up an active community of scholars who are interested in this topic.

Incidental Exposure to Online News

Rapid technological changes and availability of news anywhere and at any moment have changed how people seek out news. Increasingly, consumers no longer take deliberate actions to read the news, instead stumbling upon news online. While the emergence of serendipitous news discovery online has been recognized in the literature, there is a limited understanding about how people experience this behavior. Based on the mixed method study that investigated online news reading behavior of residents in a Midwestern U.S. town, we explore how people accidentally discover news when engaged in various online activities. Employing the grounded theory approach, we define Incidental Exposure to Online News (IEON) as individual's memorable experiences of chance encounters with interesting, useful, or surprising news while using the Internet for news browsing or for non-news-related online activities, such as checking email or visiting social networking sites. The book presents a conceptual framework of IEON that advances research and an understanding of serendipitous news discovery from people's holistic experiences of news consumption in their everyday lives. The proposed IEON Process Model identifies key steps in an IEON experience that could help news reporters and developers of online news platforms create innovative storytelling and design strategies to catch consumers' attention during their online activities. Finally, this book raises important methodological questions for further investigation: how should serendipitous news discovery be studied, measured, and observed, and what are the essential elements that differentiate this behavior from other types of online news consumption and information behaviors?

Compatibility Modeling

Nowadays, fashion has become an essential aspect of people's daily life. As each outfit usually comprises several complementary items, such as a top, bottom, shoes, and accessories, a proper outfit largely relies on the harmonious matching of these items. Nevertheless, not everyone is good at outfit composition, especially those who have a poor fashion aesthetic. Fortunately, in recent years the number of online fashion-oriented communities, like IQON and Chictopia, as well as e-commerce sites, like Amazon and eBay, has grown. The tremendous amount of real-world data regarding people's various fashion behaviors has opened a door to automatic clothing matching. Despite its significant value, compatibility modeling for clothing matching that assesses the compatibility score for a given set of (equal or more than two) fashion items, e.g., a blouse and a skirt, yields tough challenges: (a) the absence of comprehensive benchmark; (b) comprehensive compatibility modeling with the multi-modal feature variables is largely untapped; (c) how to utilize the domain knowledge to guide the machine learning; (d) how to enhance the interpretability of the compatibility modeling; and (e) how to model the user factor in the personalized compatibility modeling. These challenges have been largely unexplored to date. In this book, we shed light on several state-of-the-art theories on compatibility modeling. In particular, to facilitate the research, we first build three large-scale benchmark datasets from different online fashion websites, including IQON and Amazon. We then introduce a general data-driven compatibility modeling scheme based on advanced neural networks. To make use of the abundant fashion domain knowledge, i.e., clothing matching rules, we next present a novel knowledge-guided compatibility modeling framework. Thereafter, to enhance the model interpretability, we put forward a prototype-wise interpretable compatibility modeling approach. Following that, noticing the subjective aesthetics of users, we extend the general compatibility modeling to the personalized version. Moreover, we further study the real-world problem of personalized capsule wardrobe creation, aiming to generate a minimum collection of garments that is both compatible and suitable for the user. Finally, we conclude the book and present future research directions, such as the generative compatibility modeling, virtual try-on with arbitrary poses, and clothing generation.

Analysis and Visualization of Citation Networks

Citation analysis—the exploration of reference patterns in the scholarly and scientific literature—has long been applied in a number of social sciences to study research impact, knowledge flows, and knowledge networks. It has important information science applications as well, particularly in knowledge representation and in information retrieval. Recent years have seen a burgeoning interest in citation analysis to help address research, management, or information service issues such as university rankings, research evaluation, or knowledge domain visualization. This renewed and growing interest stems from significant improvements in the availability and accessibility of digital bibliographic data (both citation and full text) and of relevant computer technologies. The former provides large amounts of data and the latter the necessary tools for researchers to conduct new types of large-scale citation analysis, even without special access to special data collections. Exciting new developments are emerging this way in many aspects of citation analysis. This book critically examines both theory and practical techniques of citation network analysis and visualization, one of the two main types of citation analysis (the other being evaluative citation analysis). To set the context for its main theme, the book begins with a discussion of the foundations of citation analysis in general, including an overview of what can and what cannot be done with citation analysis (Chapter 1). An in-depth examination of the generally accepted steps and procedures for citation network analysis follows, including the concepts and techniques that are associated with each step (Chapter 2). Individual issues that are particularly important in citation network analysis are then scrutinized, namely: field delineation and data sources for citation analysis (Chapter 3); disambiguation of names and references (Chapter 4); and visualization of citation networks (Chapter 5). Sufficient technical detail is provided in each chapter so the book can serve as a practical how-to guide to conducting citation network analysis and visualization studies. While the discussion of most of the topics in this book applies to all types of citation analysis, the structure of the text and the details of procedures, examples, and tools covered here are geared to citation network analysis rather than evaluative citation analysis. This conscious choice was based on the authors' observation that, compared to evaluative citation analysis, citation network analysis has not been covered nearly as well by dedicated books, despite the fact that it has not been subject to nearly as much severe criticism and has been substantially enriched in recent years with new theory and techniques from research areas such as network science, social network analysis, or information visualization. Table of Contents: Acknowledgment / Dedications /Foundations of Citation Analysis / Conducting Citation Network Analysis: Steps, Concepts, Techniques, and Tools / Field Delineation and

Data Sources for Citation Analysis / Disambiguation in Citation Network Analysis / Visualization of Citation Networks / Appendix 3.3 / Appendix 5.4.2 / Bibliography / Author Biographies

Social Monitoring for Public Health

Public health thrives on high-quality evidence, yet acquiring meaningful data on a population remains a central challenge of public health research and practice. Social monitoring, the analysis of social media and other user-generated web data, has brought advances in the way we leverage population data to understand health. Social media offers advantages over traditional data sources, including real-time data availability, ease of access, and reduced cost. Social media allows us to ask, and answer, questions we never thought possible. This book presents an overview of the progress on uses of social monitoring to study public health over the past decade. We explain available data sources, common methods, and survey research on social monitoring in a wide range of public health areas. Our examples come from topics such as disease surveillance, behavioral medicine, and mental health, among others. We explore the limitations and concerns of these methods. Our survey of this exciting new field of data-driven research lays out future research directions.

Learning from Multiple Social Networks

With the proliferation of social network services, more and more social users, such as individuals and organizations, are simultaneously involved in multiple social networks for various purposes. In fact, multiple social networks characterize the same social users from different perspectives, and their contexts are usually consistent or complementary rather than independent. Hence, as compared to using information from a single social network, appropriate aggregation of multiple social networks offers us a better way to comprehensively understand the given social users. Learning across multiple social networks brings opportunities to new services and applications as well as new insights on user online behaviors, yet it raises tough challenges: (1) How can we map different social network accounts to the same social users? (2) How can we complete the item-wise and block-wise missing data? (3) How can we leverage the relatedness among sources to strengthen the learning performance? And (4) How can we jointly model the dual-heterogeneities: multiple tasks exist for the given application and each task has various features from multiple sources? These questions have been largely unexplored to date. We noticed this timely opportunity, and in this book we present some state-of-the-art theories and novel practical applications on aggregation of multiple social networks. In particular, we first introduce multi-source dataset construction. We then introduce how to effectively and efficiently complete the item-wise and block-wise missing data, which are caused by the inactive social users in some social networks. We next detail the proposed multi-source mono-task learning model and its application in volunteerism tendency prediction. As a counterpart, we also present a mono-source multi-task learning model and apply it to user interest inference. We seamlessly unify these models with the so-called multi-source multi-task learning, and demonstrate several application scenarios, such as occupation prediction. Finally, we conclude the book and figure out the future research directions in multiple social network learning, including the privacy issues and source complementarity modeling. This is preliminary research on learning from multiple social networks, and we hope it can inspire more active researchers to work on this exciting area. If we have seen further it is by standing on the shoulders of giants.

Automatic Disambiguation of Author Names in Bibliographic Repositories

This book deals with a hard problem that is inherent to human language: ambiguity. In particular, we focus on author name ambiguity, a type of ambiguity that exists in digital bibliographic repositories, which occurs when an author publishes works under distinct names or distinct authors publish works under similar names. This problem may be caused by a number of reasons, including the lack of standards and common practices, and the decentralized generation of bibliographic content. As a consequence, the quality of the main services of digital bibliographic repositories such as search, browsing, and recommendation may be severely affected by author name ambiguity. The focal point of the book is on automatic methods, since manual solutions do not scale to the size of the current repositories or the speed in which they are updated. Accordingly, we provide an ample view on the problem of automatic disambiguation of author names, summarizing the results of more than a decade of research on this topic conducted by our group, which were reported in more than a dozen publications that received over 900 citations so far, according to Google Scholar. We start by discussing its motivational issues (Chapter 1). Next, we formally define the author name disambiguation task

(Chapter 2) and use this formalization to provide a brief, taxonomically organized, overview of the literature on the topic (Chapter 3). We then organize, summarize and integrate the efforts of our own group on developing solutions for the problem that have historically produced state-of-the-art (by the time of their proposals) results in terms of the quality of the disambiguation results. Thus, Chapter 4 covers HHC - Heuristic-based Clustering, an author name disambiguation method that is based on two specific real-world assumptions regarding scientific authorship. Then, Chapter 5 describes SAND - Self-training Author Name Disambiguator and Chapter 6 presents two incremental author name disambiguation methods, namely INDi - Incremental Unsupervised Name Disambiguation and INC- Incremental Nearest Cluster. Finally, Chapter 7 provides an overview of recent author name disambiguation methods that address new specific approaches such as graph-based representations, alternative predefined similarity functions, visualization facilities and approaches based on artificial neural networks. The chapters are followed by three appendices that cover, respectively: (i) a pattern matching function for comparing proper names and used by some of the methods addressed in this book; (ii) a tool for generating synthetic collections of citation records for distinct experimental tasks; and (iii) a number of datasets commonly used to evaluate author name disambiguation methods. In summary, the book organizes a large body of knowledge and work in the area of author name disambiguation in the last decade, hoping to consolidate a solid basis for future developments in the field.

Measuring User Engagement

User engagement refers to the quality of the user experience that emphasizes the positive aspects of interacting with an online application and, in particular, the desire to use that application longer and repeatedly. User engagement is a key concept in the design of online applications (whether for desktop, tablet or mobile), motivated by the observation that successful applications are not just used, but are engaged with. Users invest time, attention, and emotion in their use of technology, and seek to satisfy pragmatic and hedonic needs. Measurement is critical for evaluating whether online applications are able to successfully engage users, and may inform the design of and use of applications. User engagement is a multifaceted, complex phenomenon; this gives rise to a number of potential measurement approaches. Common ways to evaluate user engagement include using self-report measures, e.g., questionnaires; observational methods, e.g. facial expression analysis, speech analysis; neuro-physiological signal processing methods, e.g., respiratory and cardiovascular accelerations and decelerations, muscle spasms; and web analytics, e.g., number of site visits, click depth. These methods represent various trade-offs in terms of the setting (laboratory versus ``in the wild''), object of measurement (user behaviour, affect or cognition) and scale of data collected. For instance, small-scale user studies are deep and rich, but limited in terms of generalizability, whereas large-scale web analytic studies are powerful but negate users' motivation and context. The focus of this book is how user engagement is currently being measured and various considerations for its measurement. Our goal is to leave readers with an appreciation of the various ways in which to measure user engagement, and their associated strengths and weaknesses. We emphasize the multifaceted nature of user engagement and the unique contextual constraints that come to bear upon attempts to measure engagement in different settings, and across different user groups and web domains. At the same time, this book advocates for the development of ``good'' measures and good measurement practices that will advance the study of user engagement and improve our understanding of this construct, which has become so vital in our wired world.

Digital Libraries Applications

Digital libraries (DLs) have evolved since their launch in 1991 into an important type of information system, with widespread application. This volume advances that trend further by describing new research and development in the DL field that builds upon the 5S (Societies, Scenarios, Spaces, Structures, Streams) framework, which is discussed in three other DL volumes in this series.While the 5S framework may be used to describe many types of information systems, and is likely to have even broader utility and appeal, we focus here on digital libraries. Drawing upon six (Akbar, Kozievitch, Leidig, Li, Murthy, Park) completed and two (Chen, Fouh) in-process dissertations, as well as the efforts of collaborating researchers, and scores of related publications, presentations, tutorials, and reports, this book demonstrates the applicability of 5S in five digital library application areas, that also have importance in the context of the WWW, Web 2.0, and innovative information systems. By integrating surveys of the state-of-the-art, newresearch, connections with formalization, case studies, and exercises/projects, this book can serve as a textbook for those interested in computing, information,

and/or library science. Chapter 1 focuses on images, explaining how they connect with information retrieval, in the context of CBIR systems. Chapter 2 gives two case studies of DLs used in education, which is one of the most common applications of digital libraries. Chapter 3 covers social networks, which are at the heart of work onWeb 2.0, explaining the construction and use of deduced graphs, that can enhance retrieval and recommendation. Chapter 4 demonstrates the value of DLs in eScience, focusing, in particular, on cyber-infrastructure for simulation. Chapter 5 surveys geospatial information in DLs, with a case study on geocoding. Given this rich content, we trust that any interested in digital libraries, or in related systems, will find this volume to be motivating, intellectually satisfying, and useful. We hope it will help move digital libraries forward into a science as well as a practice. We hope it will help build community that will address the needs of the next generation of DLs.

Social Media and Library Services

The rise of social media technologies has created new ways to seek and share information for millions of users worldwide, but also has presented new challenges for libraries in meeting users where they are within social spaces. From social networking sites such as Facebook and Google+, and microblogging platforms such as Twitter and Tumblr to the image and video sites of YouTube, Flickr, Instagram, and to geotagging sites such as Foursquare, libraries have responded by establishing footholds within a variety of social media platforms and seeking new ways of engaging with online users in social spaces. Libraries are also responding to new social review sites such as Yelp and Tripadvisor, awareness sites including StumbleUpon, Pinterest, Goodreads, and Reddit, and social question-and-answer (Q&A) sites such as Yahoo! Answers—sites which engage social media users in functions similar to traditional library content curation, readers' advisory, information and referral, and reference services. Establishing a social media presence extends the library's physical manifestation into virtual space and increases the library's visibility, reach, and impact. However, beyond simply establishing a social presence for the library, a greater challenge is building effective and engaging social media sites that successfully adapt a library's visibility, voice, and presence to the unique contexts, audiences, and cultures within diverse social media sites. This lecture examines the research and theory on social media and libraries, providing an overview of what is known and what is not yet known about libraries and social media. Chapter 1 focuses on the social media environments within which libraries are establishing a presence, including how social media sites differ from each other, yet work together within a social ecosphere. Chapter 2 examines how libraries are engaging with users across a variety of social media platforms and the extent to which libraries are involved in using these different social media platforms, as well as the activities of libraries in presenting a social "self," sharing information, and interacting with users via social media. Chapter 3 explores metrics and measures for assessing the impact of the library's activity in social media sites. The book concludes with Chapter 4 on evolving directions for libraries and social media, including potential implications of new and emerging technologies for libraries in social spaces. Table of Contents: Preface / The Social Media Environment / Libraries and Social Media / Assessing Social Media Sites and Services / Evolving Directions in Social Libraries / Bibliography / Author Biography

Digital Library Technologies

Digital libraries (DLs) have introduced new technologies, as well as leveraging, enhancing, and integrating related technologies, since the early 1990s. These efforts have been enriched through a formal approach, e.g., the 5S (Societies, Scenarios, Spaces, Structures, Streams) framework, which is discussed in two earlier volumes in this series. This volume should help advance work not only in DLs, but also in the WWW and other information systems. Drawing upon four (Kozievitch, Murthy, Park, Yang) completed and three (Elsherbiny, Farag, Srinivasan) in-process dissertations, as well as the efforts of collaborating researchers and scores of related publications, presentations, tutorials, and reports, this book should advance the DL field with regard to at least six key technologies.
By integrating surveys of the state-of-the-art, new research, connections with formalization, case studies, and exercises/projects, this book can serve as a computing or information science textbook. It can support studies in cyber-security, document management, hypertext/hypermedia, IR, knowledge management, LIS, multimedia, and machine learning. Chapter 1, with a case study on fingerprint collections, focuses on complex (composite, compound) objects, connecting DL and related work on buckets, DCC, and OAI-ORE. Chapter 2, discussing annotations, as in hypertext/hypermedia, emphasizes parts of documents, including images as well as text, managing superimposed information. The SuperIDR system, and prototype efforts with Flickr, should motivate further development and standardization related to annotation, which would benefit all DL and WWW users. Chapter 3, on

ontologies, explains how they help with browsing, query expansion, focused crawling, and classification. This chapter connects DLs with the Semantic Web, and uses CTRnet as an example. Chapter 4, on (hierarchical) classification, leverages LIS theory, as well as machine learning, and is important for DLs as well as the WWW. Chapter 5, on extraction from text, covers document segmentation, as well as how to construct a database from heterogeneous collections of references (from ETDs); i.e., converting strings to canonical forms. Chapter 6 surveys the security approaches used in information systems, and explains how those approaches can apply to digital libraries which are not fully open. Given this rich content, those interested in DLs will be able to find solutions to key problems, using the right technologies and methods. We hope this book will help show how formal approaches can enhance the development of suitable technologies and how they can be better integrated with DLs and other information systems.

Web Indicators for Research Evaluation

In recent years there has been an increasing demand for research evaluation within universities and other research-based organisations. In parallel, there has been an increasing recognition that traditional citation-based indicators are not able to reflect the societal impacts of research and are slow to appear. This has led to the creation of new indicators for different types of research impact as well as timelier indicators, mainly derived from the Web. These indicators have been called altmetrics, webometrics or just web metrics. This book describes and evaluates a range of web indicators for aspects of societal or scholarly impact, discusses the theory and practice of using and evaluating web indicators for research assessment and outlines practical strategies for obtaining many web indicators. In addition to describing impact indicators for traditional scholarly outputs, such as journal articles and monographs, it also covers indicators for videos, datasets, software and other non-standard scholarly outputs. The book describes strategies to analyse web indicators for individual publications as well as to compare the impacts of groups of publications. The practical part of the book includes descriptions of how to use the free software Webometric Analyst to gather and analyse web data. This book is written for information science undergraduate and Master's students that are learning about alternative indicators or scientometrics as well as Ph.D. students and other researchers and practitioners using indicators to help assess research impact or to study scholarly communication.

Click Models for Web Search

With the rapid growth of web search in recent years the problem of modeling its users has started to attract more and more attention of the information retrieval community. This has several motivations. By building a model of user behavior we are essentially developing a better understanding of a user, which ultimately helps us to deliver a better search experience. A model of user behavior can also be used as a predictive device for non-observed items such as document relevance, which makes it useful for improving search result ranking. Finally, in many situations experimenting with real users is just infeasible and hence user simulations based on accurate models play an essential role in understanding the implications of algorithmic changes to search engine results or presentation changes to the search engine result page. In this survey we summarize advances in modeling user click behavior on a web search engine result page. We present simple click models as well as more complex models aimed at capturing non-trivial user behavior patterns on modern search engine result pages. We discuss how these models compare to each other, what challenges they have, and what ways there are to address these challenges. We also study the problem of evaluating click models and discuss the main applications of click models.

The Taxobook

This book is the third of a three-part series on taxonomies, and covers putting your taxonomy into use in as many ways as possible to maximize retrieval for your users. Chapter 1 suggests several items to research and consider before you start your implementation and integration process. It explores the different pieces of software that you will need for your system and what features to look for in each. Chapter 2 launches with a discussion of how taxonomy terms can be used within a workflow, connecting two—or more—taxonomies, and intelligent coordination of platforms and taxonomies. Microsoft SharePoint is a widely used and popular program, and I consider their use of taxonomies in this chapter. Following that is a discussion of taxonomies and semantic integration and then the relationship between indexing and the hierarchy of a taxonomy. Chapter 3 ("How is a Taxonomy Connected to Search?") provides discussions and examples of putting taxonomies into use in practical

applications. It discusses displaying content based on search, how taxonomy is connected to search, using a taxonomy to guide a searcher, tools for search, including search engines, crawlers and spiders, and search software, the parts of a search-capable system, and then how to assemble that search-capable system. This chapter also examines how to measure quality in search, the different kinds of search, and theories on search from several famous theoreticians—two from the 18th and 19th centuries, and two contemporary. Following that is a section on inverted files, parsing, discovery, and clustering. While you probably don't need a comprehensive understanding of these concepts to build a solid, workable system, enough information is provided for the reader to see how they fit into the overall scheme. This chapter concludes with a look at faceted search and some possibilities for search interfaces. Chapter 4, "Implementing a Taxonomy in a Database or on a Website," starts where many content systems really should—with the authors, or at least the people who create the content. This chapter discusses matching up various groups of related data to form connections, data visualization and text analytics, and mobile and e-commerce applications for taxonomies. Finally, Chapter 5 presents some educated guesses about the future of knowledge organization. Table of Contents: List of Figures / Preface / Acknowledgments / On Your Mark, Get Ready .... WAIT! Things to Know Before You Start the Implementation Step / Taxonomy and Thesaurus Implementation / How is a Taxonomy Connected to Search? / Implementing a Taxonomy in a Database or on a Website / What Lies Ahead for Knowledge Organization? / Glossary / End Notes / Author Biography

Digital Libraries for Cultural Heritage

European digital libraries have existed in diverse forms and with quite different functions, priorities, and aims. However, there are some common features of European-based initiatives that are relevant to non-European communities. There are now many more challenges and changes than ever before, and the development rate of new digital libraries is ever accelerating. Delivering educational, cultural, and research resources-especially from major scientific and cultural organizations-has become a core mission of these organizations. Using these resources they will be able to investigate, educate, and elucidate, in order to promote and disseminate and to preserve civilization. Extremely important in conceptualizing the digital environment priorities in Europe was its cultural heritage and the feeling that these rich resources should be open to Europe and the global community. In this book we focus on European digitized heritage and digital culture, and its potential in the digital age. We specifically look at the EU and its approaches to digitization and digital culture, problems detected, and achievements reached, all with an emphasis on digital cultural heritage. We seek to report on important documents that were prepared on digitization; copyright and related documents; research and education in the digital libraries field under the auspices of the EU; some other European and national initiatives; and funded projects. The aim of this book is to discuss the development of digital libraries in the European context by presenting, primarily to non-European communities interested in digital libraries, the phenomena, initiatives, and developments that dominated in Europe. We describe the main projects and their outcomes, and shine a light on the number of challenges that have been inspiring new approaches, cooperative efforts, and the use of research methodology at different stages of the digital libraries development. The specific goals are reflected in the structure of the book, which can be conceived as a guide to several main topics and sub-topics. However, the author's scope is far from being comprehensive, since the field of digital libraries is very complex and digital libraries for cultural heritage is even moreso.

Framing Privacy in Digital Collections with Ethical Decision Making

As digital collections continue to grow, the underlying technologies to serve up content also continue to expand and develop. As such, new challenges are presented which continue to test ethical ideologies in everyday environs of the practitioner. There are currently no solid guidelines or overarching codes of ethics to address such issues. The digitization of modern archival collections, in particular, presents interesting conundrums when factors of privacy are weighed and reviewed in both small and mass digitization initiatives. Ethical decision making needs to be present at the onset of project planning in digital projects of all sizes, and we also need to identify the role and responsibility of the practitioner to make more virtuous decisions on behalf of those with no voice or awareness of potential privacy breaches. In this book, notions of what constitutes private information are discussed, as is the potential presence of such information in both analog and digital collections. This book lays groundwork to introduce the topic of privacy within digital collections by providing some examples from documented real-world scenarios and making recommendations for future research. A discussion of the notion privacy as concept will be included, as well as some historical perspective (with perhaps one the

most cited work on this topic, for example, Warren and Brandeis' "Right to Privacy," 1890). Concepts from the The Right to Be Forgotten case in 2014 (Google Spain SL, Google Inc. v Agencia Españla de Protección de Datos, Mario Costeja González) are discussed as to how some lessons may be drawn from the response in Europe and also how European data privacy laws have been applied. The European ideologies are contrasted with the Right to Free Speech in the First Amendment in the U.S., highlighting the complexities in setting guidelines and practices revolving around privacy issues when applied to real life scenarios. Two ethical theories are explored: Consequentialism and Deontological. Finally, ethical decision making models will also be applied to our framework of digital collections. Three case studies are presented to illustrate how privacy can be defined within digital collections in some real-world examples.

Trustworthy Communications and Complete Genealogies

Genealogies document relationships between persons involved in historical events. Information about the events is parsed from communications from the past. This book explores a way to organize information from multiple communications into a trustworthy representation of a genealogical history of the modern world. The approach defines metrics for evaluating the consistency, correctness, closure, connectivity, completeness, and coherence of a genealogy. The metrics are evaluated using a 312,000-person research genealogy that explores the common ancestors of the royal families of Europe. A major result is that completeness is defined by a genealogy symmetry property driven by two exponential processes, the doubling of the number of potential ancestors each generation, and the rapid growth of lineage coalescence when the number of potential ancestors exceeds the available population. A genealogy expands from an initial root person to a large number of lineages, which then coalesce into a small number of progenitors. Using the research genealogy, candidate progenitors for persons of Western European descent are identified. A unifying ancestry is defined to which historically notable persons can be linked.

Mobile Search Behaviors

With the rapid development of mobile Internet and smart personal devices in recent years, mobile search has gradually emerged as a key method with which users seek online information. In addition, cross-device search also has been regarded recently as an important research topic. As more mobile applications (APPs) integrate search functions, a user's mobile search behavior on different APPs becomes more significant. This book provides a systematic review of current mobile search analysis and studies user mobile search behavior from several perspectives, including mobile search context, APP usage, and different devices. Two different user experiments to collect user behavior data were conducted. Then, through the data from user mobile phone usage logs in natural settings, we analyze the mobile search strategies employed and offer a context-based mobile search task collection, which then can be used to evaluate the mobile search engine. In addition, we combine mobile search with APP usage to give more in-depth analysis, such as APP transition in mobile search and follow-up actions triggered by mobile search. The study, combining the mobile search with APP usage, can contribute to the interaction design of APPs, such as the search recommendation and APP recommendation. Addressing the phenomenon of users owning more smart devices today than ever before, we focus on user cross device search behavior. We model the information preparation behavior and information resumption behavior in cross-device search and evaluate the search performance in cross-device search. Research on mobile search behaviors across different devices can help to understand online user information behavior comprehensively and help users resume their search tasks on different devices.

Task Intelligence for Search and Recommendation

While great strides have been made in the field of search and recommendation, there are still challenges and opportunities to address information access issues that involve solving tasks and accomplishing goals for a wide variety of users. Specifically, we lack intelligent systems that can detect not only the request an individual is making (what), but also understand and utilize the intention (why) and strategies (how) while providing information and enabling task completion. Many scholars in the fields of information retrieval, recommender systems, productivity (especially in task management and time management), and artificial intelligence have recognized the importance of extracting and understanding people's tasks and the intentions behind performing those tasks in order to serve them better. However, we are still struggling to support them in task completion, e.g., in search and

assistance, and it has been challenging to move beyond single-query or single-turn interactions. The proliferation of intelligent agents has unlocked new modalities for interacting with information, but these agents will need to be able to work understanding current and future contexts and assist users at task level. This book will focus on task intelligence in the context of search and recommendation. Chapter 1 introduces readers to the issues of detecting, understanding, and using task and task-related information in an information episode (with or without active searching). This is followed by presenting several prominent ideas and frameworks about how tasks are conceptualized and represented in Chapter 2. In Chapter 3, the narrative moves to showing how task type relates to user behaviors and search intentions. A task can be explicitly expressed in some cases, such as in a to-do application, but often it is unexpressed. Chapter 4 covers these two scenarios with several related works and case studies. Chapter 5 shows how task knowledge and task models can contribute to addressing emerging retrieval and recommendation problems. Chapter 6 covers evaluation methodologies and metrics for task-based systems, with relevant case studies to demonstrate their uses. Finally, the book concludes in Chapter 7, with ideas for future directions in this important research area.

Word Association Thematic Analysis

Many research projects involve analyzing sets of texts from the social web or elsewhere to get insights into issues, opinions, interests, news discussions, or communication styles. For example, many studies have investigated reactions to Covid-19 social distancing restrictions, conspiracy theories, and anti-vaccine sentiment on social media. This book describes word association thematic analysis, a mixed methods strategy to identify themes within a collection of social web or other texts. It identifies these themes in the differences between subsets of the texts, including female vs. male vs. nonbinary, older vs. newer, country A vs. country B, positive vs. negative sentiment, high scoring vs. low scoring, or subtopic A vs. subtopic B. It can also be used to identify the differences between a topic-focused collection of texts and a reference collection. The method starts by automatically finding words that are statistically significantly more common in one subset than another, then identifies the context of these words and groups them into themes. It is supported by the free Windows-based software Mozdeh for data collection or importing and for the quantitative analysis stages. This book explains the word association thematic analysis method, with examples, and gives practical advice for using it. It is primarily intended for social media researchers and students, although the method is applicable to any collection of short texts.